

Chapter 5: Data Quality

This section provides detailed information on the lineage, positional accuracy, attribute accuracy, logical consistency, and completeness of the 1995 TIGER/Line® files. Data users can use this information to help evaluate the adequacy and applicability of this geographic file for a particular use.

Lineage

Geometric Properties

Source codes that specify the original digital source of complete chains in the TIGER/Line® files are listed in the *Sources* section of this chapter. These codes cover the source categories in the Census TIGER® data base: initial source, pre-1990 computer operations, office operations, enumerator operations, local official updates, post-1990 census updates, and pre-2000 computer operations.

The initial sources used to create the Census TIGER® data base were the USGS 1:100,000-scale Digital Line Graph (DLG), USGS 1:24,000-scale quadrangles, the Census Bureau's 1980 geographic base files (GBF/DIME-Files), and a variety of miscellaneous maps for selected areas outside the contiguous 48 states. The DLG coverage is extensive, albeit of variable currency, and comprises most of the rural, small city, and suburban area of the TIGER/Line® files. GBF/DIME-File coverage areas were updated through 1987 with the manual translation of features from the most recent aerial photography available to the Census Bureau.

In order to create a current geographic data base for the extraction of the 1995 TIGER/Line® files, the Census Bureau used various internal and external procedures through 1995 to update and maintain the Census TIGER® data base. While the Census Bureau has made a reasonable and systematic attempt to gather the most recent information available about the features that this file portrays, the Census Bureau cautions users that the files are no more complete than the source documents used in their compilation, the vintage of those source documents, and the translation of the information on those source documents.

The Census Bureau added, to the Census TIGER® data base, the enumerator updates compiled during the 1988-1990 census operations. The updates came from map annotations made by enumerators as they attempted to locate living quarters by traversing every street feature in their assignment area. The Census Bureau digitized the enumerator updates directly into the Census TIGER® data base without geodetic controls or the use of aerial photography to confirm the features' existence or locational accuracy.

The Census Bureau also made other corrections and updates to the map sheets supplied by local participants in various Census Bureau programs. Unconfirmed local updates originated from map reviews by local government officials or their liaisons. Maps were sent to the highest elected official of governmental units between 1990-1995 for use in various census programs, and some maps were returned with update annotations and corrections. The Census Bureau generally added the updates to the Census TIGER® data base without extensive checks when the elected official approved the boundary or feature correction. Changes made by local officials do not have geodetic control.

Projection

The TIGER/Line® data is not in a mapping projection even though most of the features were scanned directly from source maps (usually USGS 1:100,000 topographic quads) that were projections. The USGS source maps were Universal Transverse Mercator (UTM) projections. After the map sheets were scanned, the coordinates were transformed from UTM into projectionless geographic coordinates of latitude and longitude. The USGS Digital Line Graphs (DLGs) were derived from the same operation, but typically were distributed as UTM projections.

As mentioned earlier, there were a variety of other sources used in creating the Census TIGER® data base. The features from those sources also were stored as latitude and longitude coordinates. For metropolitan areas, the GBF/DIME files were derived by digitizing a variety of sources, (in various projections) such as USGS topographic 7.5 minute quadrangles, aerial photography, and other materials, to fill in the gaps between the GBF/DIME coverage and the 1:100,000 topographic sheet coverage. Also

included in the Census TIGER® data base are features obtained from field updates. Paper maps were annotated in the field and subsequently digitized without rigorous adherence to a projection or coordinate system.

Sources

In the TIGER/Line® files, there is a 1-alphanumeric character source code for complete chain and landmark features. Source codes identify the original (or final, if historical) operation that created the geographic object and its geometric properties.

Source Codes

<i>Value</i>	<i>Description</i>
<i>blank</i>	Not Documented Elsewhere
A	Updated 1980 GBF/DIME-File
B	USGS 1:100,000-Scale DLG-3 File
C	Other USGS Map
D	Census Bureau Update Prior to 1990 Enumeration Operations
E	Census Bureau 1990 Enumerator Update
F	Census Bureau Update from Other 1990 Operations
G	Unconfirmed Local Official Updates
H	Census Bureau Update Post-1990 Operations
I	Census Address List/TIGER Linkage Operations

Source Code Record Locations

<i>Record Type</i>	<i>Field Name</i>	<i>Description</i>
1	SOURCE	Source or First Source Code of Update
7	SOURCE	Source or First Source Code to Update
9	SOURCE	Source or First Source Code to Update
H	HIST	History or Last Source Code to Update
H	SOURCE	Source or First Source Code to Update

Address Ranges and ZIP Codes®

The TIGER/Line® files contain potential address ranges and ZIP Codes® for most areas of the United States where city-style address ranges exist. Residential addresses from the 1990 decennial census master list of addresses, the Address Control File (ACF), were converted to address

ranges and matched into TIGER using an address range creation formula for all counties. The original TIGER address ranges were matched, then merged with the ACF-derived address ranges, producing a single set of integrated address ranges in the TIGER data base. Thus, the 1992 TIGER/Line® files contained a mixture of pre-existing TIGER address ranges from these areas, or ACF ranges where no other range was available. No attempt was made to reconcile any overlapping address ranges or close any coverage gaps in the 1992 TIGER/Line® files.

Subsequently, during the ACF Match/Merge operation, the ranges were integrated and many address range conflicts were resolved. Further address range edits eliminated or isolated additional overlaps. The edited address ranges appear in the 1995 TIGER/Line® files.

ZIP Codes® were originally derived from two sources: those already existing in the Census TIGER® data base and those derived from the ACF. Address ranges created from the ACF may have non-city delivery ZIP Codes®. This situation typically occurs in smaller places where structure numbers exist and appear in the ACF, but are not used in mail delivery.

The ZIP Codes® were updated and corrected by matching the Census TIGER® data base with an updated USPS ZIP+4® file (AMS Match) for the 50 states and the District of Columbia. The 5-digit ZIP Code® and street name were used as keys to match address ranges from the TIGER data base to corresponding address ranges in the ZIP+4® file. Where a match occurred, the ZIP Add-On (Plus 4) code was added to the TIGER address range record. If the TIGER data base address ranges did not align exactly with the ZIP+4® address ranges, the TIGER ranges were split into multiple records. This procedure allowed the Add-On code to transfer more easily. Two rounds of matches have been completed. The first used data from the August 1993 ZIP+4® file, and the latest used data from the July 1994 file. Between the matches, clerical updates improved five-digit ZIP Code® coverage (particularly notable in southern California and central Florida), and eliminated the illegal five-digit ZIP Codes® and three-digit ZIP Codes®. Many of the ZIP Code® changes reflect regular realignments and the creation of new ZIP Codes® in high growth areas of the country.

Census Feature Class Codes

All generic CFCCs (A10, A20, A30, and A40) were changed to more descriptive CFCCs. For example, an A40 (local, neighborhood, and rural road, major category used alone when the minor category could not be determined) was changed to the more descriptive CFCC of A41 (unseparated local, neighborhood and rural road). The census feature classifications of roads were redefined to agree more closely with customary use and to be more useful to transportation planners. Thus, all road classifications were reduced to a local or neighborhood road unless the road had a highway route number. The classification was then based on the highway route number.

Feature Identifiers

Highway Route Numbers The Census Bureau updated the feature identifiers (FIDs) and census feature class codes (CFCCs) for all interstates, limited access roads, US highways, and state highways in all counties in the United States. The FIDs of highways were entered in the Census TIGER® data base using the following rules:

- If an interstate also was known by a local name, the interstate route number was entered as the primary name of the interstate and the local name was entered as the alternate name.
- If the US highways and state highways were known by a route number as well as by a local name, the local name was entered as the primary name, and the highway route number was entered as the alternate name.

Railroad Names The Census Bureau has been working to complete an ongoing project to update the railroad names in the Census TIGER® data base. National update of the railroad names has not yet been completed.

Positional Accuracy

The Census Bureau's mission to count and profile the Nation's people and institutions does not require very high levels of positional accuracy in its geographic products. Its files and maps are designed to show only the relative positions of elements.

Coordinates in the TIGER/Line[®] files have six implied decimal places. The positional accuracy of these coordinates is not as great as the six decimal places suggest. The positional accuracy varies with the source materials used, but at best meets the established National Map Accuracy standards (approximately ± 167 feet) where 1:100,000-scale maps from the USGS are the source. The Census Bureau can not specify the accuracy of feature updates added by its field staff or of features derived from the GBF/DIME-Files or other map sources. Thus, the level of positional accuracy in the 1995 TIGER/Line[®] files is not suitable for high-precision measurement applications such as engineering problems, property transfers, or other uses that might require highly accurate measurements of the earth's surface.

Despite the fact that TIGER/Line[®] data positional accuracy is not as high as the coordinate values imply, the six-decimal place precision is useful when producing maps. This precision allows you to place features that are next to each other on the ground in the correct position, relative to each other, on the map without overlap.

Attribute Accuracy

Topological Properties

The attribute accuracy of the TIGER/Line[®] files is as precise as the source used during the creation or update of the Census TIGER[®] data base. Accuracy statements on the Census TIGER[®] data base are based on deductive estimates; no specific field tests for attribute accuracy have been conducted on the files. However, updates or corrections resulting from normal Census Bureau field operations are entered into the Census TIGER[®] data base. In addition, quality checks are conducted to verify clerical transcription of data from source materials. Based on past experience, attribute codes match the source materials with less than a two-percent error.

The feature network of complete chains (as represented by Record Types 1 and 2) is complete for census purposes. Data users should be aware that on occasion they may not be able to trace a specific feature by name or by CFCC as a continuous line throughout the TIGER/Line[®] files without making additional edits. For example, State Highway 32

may cross the entire county. The TIGER/Line® files will contain complete chains in the file at the location of State Highway 32, but the complete chains may individually have one of a collection of local names such as S Elm Street, or Smallville Highway, with or without State Highway 32 as an alternate. The most frequent CFCC for a state highway is A21, but the complete chains at the location of State Highway 32 may have a variety of class codes such as A01, A41, or A21. Recent edits have reduced this problem, but not eliminated it.

Boundaries and Geographic Entity Codes

The Census Bureau collects and tabulates information for both legal and statistical entities. Record Types 1 and S identify the boundaries and codes for the legal entities reported to the Census Bureau to be legally in effect on January 1, 1995. Record Types 3 and A generally contain the final 1990 census tabulation geographic boundaries and codes for those entities. Most legal boundaries are based on the annotations made by local officials in response to the Census Bureau's Boundary and Annexation Surveys.

Census Statistical Area Committees (CSACs) generally define and delineate statistical entities following Census Bureau guidelines. However, there are several exceptions:

- UAs are defined strictly by the Census Bureau based on technical considerations.
- School districts are delineated by State Departments of Education.
- Voting Districts (VTDs) are supplied by the designated liaison for the 1990 Redistricting Data Program.

The USGS maintains the file that is published as FIPS 55. The Census Bureau uses the file for coding American Indian/Alaska Native Areas, county subdivisions, consolidated cities, places, and sub-MCDs. Cooperatively in 1993, the Census Bureau and the USGS edited the FIPS 55 file to ensure alphabetical sorting and data consistency. As a result, several hundred changes were made to the FIPS 55 codes and related class codes. These corrections, plus codes for new entities, appear in Record Type C where one record shows the codes used in 1990, and one record shows the current FIPS 55 code/class code associated with each change.

Other attribute data in the TIGER/Line® files were gathered from many sources. The Census Bureau's staff linked the attribute information to the spatial framework of features. Most procedures for gathering the needed attributes were clerical. The quality of these attributes was ensured by various tests conducted before, during, and after the time that the attribute information was entered into the Census TIGER® data base. Tests included source material selection and evaluation checks, quality control checks on staff work, independent reviews by local and tribal leaders of maps produced from the Census TIGER® data base, and staff reviews of computer-performed operations.

Address Ranges and ZIP Codes®

The conversion from the GBF/DIME-Files to the TIGER format involved neither verification of previously existing address ranges nor any significant updates or corrections (except as noted below). Prior to the release of the 1992 TIGER/Line® files, the address ranges for an area were generally the same as those in the corresponding 1980 GBF/DIME-File. Preparations for the 1990 census involved making some minor updates in selected areas, but generally did not include changes in address numbering systems during the decade. The 1992 TIGER/Line® files included ACF address ranges for existing and new features identified during census operations. Users of the 1992 TIGER/Line® file's address ranges need to check for address range problems such as overlaps, gaps, odd/even reversals, and other situations that may exist. Users of the 1995 TIGER/Line® files have the benefit of the files going through a TIGER/ACF Match and Merge operation and subsequent address range edits. Many of the previous problems with addresses have been corrected.

Although an address range in the TIGER/Line® files may be incorrect, the Census Bureau implemented procedures to ensure that the error did not adversely affect the accuracy or the quality of the 1990 census. For the geographic areas with the GBF/DIME-File and extension area coverage, the Census Bureau used the address ranges to perform an initial assignment of residential addresses (purchased from a commercial vendor) to the 1990 census tract and block numbers, and made a

number of corrections to the address ranges. Later during field operations, enumerators updated, verified, and corrected, when necessary, the addresses assigned to each block number by walking the perimeter and all interior streets of each census block.

Address ranges and ZIP Codes® are being verified and coverage extended for the 2000 census through the use of the Master Address File (MAF). The MAF will be closely linked to the Census TIGER® data base. Local address lists and address ranges from the US Postal Service will supplement the MAF.

Feature Identifiers

A national consistency review of all feature names in the Census TIGER® data base was performed by running a revised name standardizer on all feature identifiers. An additional benefit was the removal of nonstandard characters and punctuation from the names. To improve accuracy, road names in the Census TIGER® data base were compared with street names in the ZIP+4® (AMS) file from the US Postal Service. Errors in feature directionals or feature types were corrected in the Census TIGER® data base.

Logical Consistency

Node-line-area relationships satisfy topological requirements. These requirements include the following:

- Complete chains must begin and end at nodes.
- Complete chains must connect to each other at nodes.
- Complete chains do not extend through nodes.
- Left and right polygons are defined for each complete chain element and are consistent for complete chains connecting at nodes.
- Complete chains representing the limits of a file are free from gaps.

The Census Bureau performed automated tests to ensure logical consistency and limits of file. Some polygons in the TIGER/Line® files may be so small, the polygon internal point has been manually placed on a node that defines the polygon perimeter. These small polygons have been detected, and corrections will be incorporated in the Census TIGER® data base in the future. The Census Bureau uses its internally developed

Geographic Update System to enhance and modify spatial and attribute data in the Census TIGER® data base.

The Census TIGER® data base has three generations of currency in geographic areas. These are generally the previous census areas, current areas, and the next census areas. The boundaries of geographic areas are affected by the location, type, and number of areas.

To prepare for the 2000 census, those features used only as boundaries in the 1980 census were deleted. The deletions lowered the overall count of complete chains and polygons.

Standard geographic codes, such as FIPS codes for states, counties, municipalities, and places, are used when encoding spatial entities. The Census Bureau performed spatial data tests for logical consistency of the codes during the compilation of the original Census TIGER® data base files. Most of the codes themselves were provided to the Census Bureau by the USGS, the agency responsible for maintaining FIPS 55.

Completeness

The GBF/DIME-Files and the USGS's DLG were the two main sources of spatial attribute data. Data for a given category contain attribute codes that reflect the information portrayed on the original source.

The TIGER/Line® files also use the Census Bureau's internal coding scheme which in some cases parallels the FIPS codes. The feature network of complete chains is complete for census purposes. For the 1990 census, census enumerators identified new and previously unreported street features for the entire Nation during a series of decennial census operations. In some areas, local officials reviewed the census maps and identified new features and feature changes. Similar operations are planned for the 2000 census. The TIGER/Line® files contain limited point and area landmark data. The enumerator updates from the 1990 census did not stress landmark features. Computer file matching and automated updates from the Economic and Agriculture censuses added landmarks and key geographic locations (KGLs). As source material from the US Postal Service and local agencies is integrated into the Census TIGER® data base, the number of landmarks and KGLs is expected to increase.